

(19)



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11)

EP 0 258 017 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:
04.06.1997 Bulletin 1997/23

(51) Int Cl.⁶: **C12N 15/10, C12N 9/12,
C12P 19/34, C12N 9/96,
C12Q 1/68**

(21) Application number: **87307433.0**

(22) Date of filing: **21.08.1987**

(54) **Purified thermostable enzyme and process for amplifying, detecting, and/or cloning nucleic acid sequences using said enzyme**

Gereinigtes thermostabiles Enzym und Verfahren zur Amplifikation, zum Nachweis und/oder zur Clonierung von Nukleinsäuresequenzen unter Verwendung dieses Enzyms

Enzyme thermostable purifiée et procédé d'amplification, de détection et/ou de clonage de séquences d'acide nucléique à l'aide de cette enzyme

(84) Designated Contracting States:
AT BE CH DE ES FR GB GR IT LI LU NL SE

(30) Priority: **22.08.1986 US 899513
22.08.1986 US 899241
17.06.1987 US 63647
17.06.1987 US 63509**

(43) Date of publication of application:
02.03.1988 Bulletin 1988/09

(60) Divisional application: **96119077.4**

(73) Proprietor: **F. HOFFMANN-LA ROCHE AG
4002 Basel (CH)**

(72) Inventors:

- Erlich, Henry Anthony
Oakland California 94602 (US)
- Horn, Glenn
Emeryville California 94608 (US)
- Saiki, Randall Keichi
Richmond California 94805 (US)
- Stoffel, Susanne
El Cerrito California 94530 (US)
- Mullis, Kary Banks
La Jolla California 92037 (US)
- Lawyer, Frances Cook
Oakland California 94611 (US)

- Gelfand, David Harrow
Oakland, California 94611 (US)

(74) Representative: **VOSSIUS & PARTNER**
Postfach 86 07 67
81634 München (DE)

(56) References cited:

- | | |
|-----------------|-----------------|
| EP-A- 0 156 287 | EP-A- 0 164 054 |
| EP-A- 0 184 086 | EP-A- 0 200 362 |
| EP-A- 0 229 701 | EP-A- 0 237 362 |
- BIOKHIMIYA, vol. 45, no. 4, April 1980, pages 644-651, Moscow, USSR; A.S. KALÉDIN et al.: "Isolation and Properties of dna Polymerase from Extremely Thermophilic Bacterium *Thermus aquaticus* YT 1"
 - JOURNAL OF BACTERIOLOGY, vol. 127, no. 3, September 1976, pages 1550-1557, Baltimore, USA; ALICE CHIEN et al.: "Deoxyribonucleic Acid Polymerase from the Extreme Thermophile *Thermus aquaticus*"
 - SCIENCE, vol. 230, no. 4732, 20 December 1985, pages 1350-1354, Washington DC, US; RANDALL K. SAIKI et al.: "Enzymatic Amplification of beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia"
 - PATENT ABSTRACTS OF JAPAN, vo. 5, no. 67 (C-53) (739), 7 May 1981; & JP-A- 5618597

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 0 258 017 B1

Description

The present invention relates to a purified thermostable enzyme having DNA polymerase activity and a molecular weight of about 86,000-90,000.

5 The present invention relates to a process for amplifying existing nucleic acid sequences if they are present in a test sample and detecting them if present by using a probe. More specifically, it relates to a process for producing any particular nucleic acid sequence from a given sequence of DNA or RNA in amounts which are large compared to the amount initially present so as to facilitate detection of the sequences, using a thermostable enzyme to catalyze the reaction. The DNA or RNA may be single- or double-stranded, and may be a relatively pure species or a component of a mixture of nucleic acids. The process of the invention utilizes a repetitive reaction to accomplish the amplification of the desired nucleic acid sequence.

Extensive research has been conducted on the isolation of DNA polymerases from mesophilic microorganisms such as *E. coli*. See, for example, Bessman et al., *J. Biol. Chem.* (1957) 233:171-177 and Buttin and Kornberg (1966) *J. Biol. Chem.* 241:5419-5427.

15 In contrast, relatively investigation investigation been been made on the isolation and purification of DNA polymerases from thermophiles, such as *Thermus aquaticus*. Kaledin et al., *Biokhimiya* (1980) 45:644-651 discloses a six-step isolation and purification procedure of DNA polymerase from cells of *T. aquaticus* YT1 strain. These steps involve isolation of crude extract, DEAE-cellulose chromatography, fractionation on hydroxyapatite, fractionation on DEAE-cellulose, and chromatography on single-strand DNA-cellulose. The pools from each stage were not screened for contaminating endo- and exonuclease(s). The molecular weight of the purified enzyme is reported as 62,000 daltons per monomeric unit.

A second purification scheme for a polymerase from *T. aquaticus* is described by A. Chien et al., *J. Bacteriol.* (1976) 127:1550-1557. In this process, the crude extract is applied to a DEAE-Sephadex column. The dialyzed pooled fractions are then subjected to treatment on a phosphocellulose column. The pooled fractions are dialyzed and bovine serum albumin (BSA) is added to prevent loss of polymerase activity. The resulting mixture is loaded on a DNA-cellulose column. The pooled material from the column is dialyzed and analyzed by gel filtration to have a molecular weight of about 63,000 daltons, and, by sucrose gradient centrifugation of about 68,000 daltons.

25 The use of a thermostable enzyme to amplify existing nucleic acid sequences in amounts that are large compared to the amount initially present has been suggested in European Pat. Pub. No. 200,362 published December 10, 1986. Primers, nucleotide triphosphates, and a polymerase are used in the process, which involves denaturation, synthesis of template strands and hybridization. The extension product of each primer becomes a template for the production of the desired nucleic acid sequence. The application discloses that if the polymerase employed is a thermostable enzyme, it need not be added after every denaturation step, because the heat will not destroy its activity. No other advantages or details are provided on the use of a purified thermostable DNA polymerase. The amplification and detection process is also described by Saiki et al., *Science*, 230:1350-1354 (1985), and by Saiki et al., *Bio/Technology*, 3:1008-1012 (1985).

Accordingly, there is a desire in the art to produce a purified, stable thermostable enzyme that may be used to improve the diagnostic amplification process described above.

40 Accordingly, the present invention, provides purified thermostable enzymes that have DNA polymerase activity, catalyze the combination of nucleoside triphosphates to form a nucleic acid strand complementary to a nucleic acid template strand and that have a molecular weight of 86,000 to 90,000 as determined according to their migration in SDS PAGE, when the molar proteins are phosphorylase B (92,500), bovine serum albumin (66,200), ovalbumin (95,000), carbonic anhydrase (31,000), soybean trypsin inhibitor (21,500) and, lysozyme (14,400). In another embodiment said thermostable enzyme is a recombinant enzyme or modification thereof. Preferably the purified enzyme is DNA polymerase from *Thermus aquaticus* and has a molecular weight of about 86,000-90,000 daltons. This purified material may be used in a temperature-cycling amplification reaction wherein nucleic acid sequences are produced from a given nucleic acid sequence in amounts that are large compared to the amount initially present so that they can be detected easily.

50 The gene encoding the enzyme from DNA polymerase from *Thermus aquaticus* has also been identified and provides yet another means to retrieve the thermostable enzyme of the present invention. In addition to the gene encoding the approximately 86-000-90,000 dalton enzyme, gene derivatives encoding DNA polymerase activity are also presented.

The invention also encompasses a stable enzyme composition comprising a purified, thermostable enzyme as described above in a buffer containing one or more non-ionic polymeric detergents.

55 The purified enzyme, as well as the enzymes produced by recombinant DNA techniques, provide much more specificity than the Klenow fragment, which is not thermostable. In addition, the purified enzyme and the recombinantly produced enzymes exhibit the appropriate activity expected when dTTP or other nucleotide triphosphates are not present in the incubation mixture with the DNA template. Also, the enzymes herein have a broader pH profile than that

of the thermostable enzyme from *Thermus aquaticus* described in the literature, with more than 50% of the activity at pH 7 as at pH 8. In addition, the thermostable enzyme herein can be stored in a buffer with non-ionic detergents so that it is stable, not losing activity over a period of time.

The present invention resides in a process for amplifying one or more specific nucleic acid sequences present in a nucleic acid or mixture thereof using primers and a thermostable enzyme. The extension product of one primer when hybridized to the other becomes a template for the production of the desired specific nucleic acid sequence, and vice versa, and the process is repeated as often as is necessary to produce the desired amount of the sequence. The method herein improves the specificity of the amplification reaction, resulting in a very distinct signal of amplified nucleic acid. In addition, the method herein eliminates the need for transferring reagents from one vessel to another after each amplification cycle. Such transferring is not required because the thermostable enzyme will withstand the high temperatures required to denature the nucleic acid strands and therefore does not need replacement. The temperature cycling may, in addition, be automated for further reduction in manpower and steps required to effectuate the amplification reaction.

More specifically, the present invention provides a process for amplifying at least one specific nucleic acid sequence contained in a nucleic acid or a mixture of nucleic acids, wherein if the nucleic acid is double-stranded, it consists of two separated complementary strands of equal or unequal length, which process comprises:

(a) contacting each nucleic acid strand with four different nucleotide triphosphates and one oligonucleotide primer for each different specific sequence being amplified, wherein each primer is selected to be substantially complementary to different strands of each specific sequence, such that the extension product synthesized from one primer, when it is separated from its complement, can serve as a template for synthesis of the extension product of the other primer, said contacting being at a temperature which promotes hybridization of each primer to its complementary nucleic acid strand;

(b) contacting each nucleic acid strand, at the same time as or after step (a), with a thermostable enzyme which enables combination of the nucleotide triphosphates to form primer extension products complementary to each strand of each nucleic acid;

(c) maintaining the mixture from step (b) at an effective temperature for an effective time to activate the enzyme, and to synthesize, for each different sequence being amplified, an extension product of each primer which is complementary to each nucleic acid strand template, but not so high (a temperature) as to separate each extension product from its complementary strand template;

(d) heating the mixture from step (c) for an effective time and at an effective temperature to separate the primer extension products from the templates on which they were synthesized to produce single-stranded molecules, but not so high (a temperature) as to denature irreversibly the enzyme;

(e) cooling the mixture from step (d) at an effective temperature for an effective time to promote hybridization of each primer to each of the single-stranded molecules produced in step (d); and

(f) maintaining the mixture from step (e) at an effective temperature for an effective time to promote the activity of the enzyme and to synthesize, for each different sequence being amplified, an extension product of each primer which is complementary to each nucleic acid strand template produced in step (d), but not so high (a temperature) as to separate each extension product from its complementary strand template, wherein steps (e) and (f) are carried out simultaneously or sequentially.

The steps (d), (e) and (f) may be repeated until the desired level of sequence amplification is obtained. The preferred thermostable enzyme herein is a polymerase extracted from *Thermus aquaticus* (Taq polymerase). Most preferably, if the enzyme is Taq polymerase, in step (a) the nucleic acid strands are contacted with a buffer comprising about 1.5-2 mM of a magnesium salt, 150-200 μ M each of the nucleotides, and 1 μ M of each primer, steps (a), (e) and (f) are carried out at about 45-58°C, and step (d) is carried out at about 90-100°C.

In a preferred embodiment, the nucleic acid(s) are double-stranded and step (a) is accomplished by (i) heating each nucleic acid in the presence of four different nucleotide triphosphates and one oligonucleotide primer for each different specific sequence being amplified, for an effective time and at an effective temperature to denature each nucleic acid, wherein each primer is selected to be substantially complementary to different strands of each specific sequence, such that the extension product synthesized from one primer, when it is separated from its complement, can serve as a template for synthesis of the extension product of the other primer; and (ii) cooling the denatured nucleic acids to a temperature which promotes hybridization of each primer to its complementary nucleic acid strand.

In other embodiments the invention relates to a process for detecting the presence or absence of at least one specific nucleic acid sequence in a sample containing a nucleic acid or mixture of nucleic acids, or distinguishing between two different sequences in said sample, wherein the sample is suspected of containing said sequence or sequences, and wherein if the nucleic acid(s) are double-stranded, they each consist of two separated complementary strands of equal or unequal length, which process comprises steps (a) to (f) mentioned above, resulting in amplification in quantity of the specific nucleic acid sequence(s), if present;

(g) adding to the product of step (f) a labeled oligonucleotide probe, for each sequence being detected, capable of hybridizing to said sequence or to a mutation thereof; and

(h) determining whether said hybridization has occurred.

In yet another embodiment, the invention relates to a process for detecting the presence or absence of at least one nucleotide variation in sequence in one or more nucleic acids contained in a sample, wherein if the nucleic acid is double-stranded it consists of two separated complementary strands of equal or unequal length, which process comprises steps (a)-(f) mentioned above, wherein steps (d), (e) and (f) are repeated a sufficient number of times to result in detectable amplification of the nucleic acid containing the sequence, if present;

(g) affixing the product of step (f) to a membrane;

(h) treating the membrane under hybridization conditions with a labeled sequence-specific oligonucleotide probe capable of hybridizing with the amplified nucleic acid sequence only if a sequence of the probe is complementary to a region of the amplified sequence; and

(i) detecting whether the probe has hybridized to an amplified sequence in the nucleic acid sample.

If the sample comprises cells, preferably they are heated before step (a) to expose the nucleic acids therein to the reagents. This step avoids extraction of the nucleic acids prior to reagent addition.

In a variation of this process, the primer(s) and/or nucleotide triphosphates are labeled so that the resulting amplified sequence is labeled. The labeled primer(s) and/or nucleotide triphosphate(s) can be present in the reaction mixture initially or added during a later cycle. The sequence-specific oligonucleotide (unlabeled) is affixed to a membrane and treated under hybridization conditions with the labeled amplification product so that hybridization will occur only if the membrane-bound sequence is present in the amplification product.

In yet another embodiment, the invention herein relates to a process for cloning into a cloning vector one or more specific nucleic acid sequences contained in a nucleic acid or a mixture of nucleic acids, which nucleic acid(s) when double-stranded consist of two separated complementary strands, and which nucleic acid(s) are amplified in quantity before cloning, which process comprises steps (a)-(f) mentioned above, with steps (d), (e) and (f) being repeated a sufficient number of times to result in detectable amplification of the nucleic acid(s) containing the sequence(s);

(g) adding to the product of step (f) a restriction enzyme for each of said restriction sites to obtain cleaved products in a restriction digest; and

(h) ligating the cleaved product(s) of step (g) containing the specific sequence(s) to be cloned into one or more cloning vectors containing a promoter and a selectable marker.

In a further embodiment, the invention herein relates to a process for cloning into a cloning vector one or more specific nucleic acid sequences contained in a nucleic acid or mixture of nucleic acids, which nucleic acid(s), when double-stranded, consist of two separated complementary strands of equal or unequal length which nucleic acid(s) are amplified in quantity before cloning, which process comprises steps (a)-(f) mentioned above, with steps (d), (e) and (f) being repeated a sufficient number of times to result in effective amplification of the nucleic acid(s) containing the sequence(s) for blunt-end ligation into one or more cloning vectors; and

(g) ligating the amplified specific sequence(s) to be cloned obtained from step (f) into one or more of said cloning vectors in the presence of a ligase, said amplified sequence(s) and vector(s) being present in sufficient amounts to effect the ligation.

In a product embodiment, the invention provides a composition of matter useful in amplifying at least one specific nucleic acid sequence contained in a nucleic acid or a mixture of nucleic acids, comprising four different nucleotide triphosphates and one oligonucleotide primer for each different specific sequence being amplified, wherein each primer is selected to be substantially complementary to different strands of each specific sequence, such that the extension product synthesized from one primer, when it is separated from its complement, can serve as a template for synthesis

of the extension product of the other primer.

In another product embodiment, the invention provides a sample of one or more nucleic acids comprising multiple strands of a specific nucleic acid sequence contained in the nucleic acid(s). The sample may comprise about 10-100 of the strands, about 100-1000 of the strands, or over about 1000 of the strands.

In a further product embodiment, the invention provides an amplified nucleic acid sequence from a nucleic acid or mixture of nucleic acids comprising multiple copies of the sequence produced by the amplification processes herein.

Figure 1 is a restriction site map of plasmid pFC83 that contains the ~4.5 kb HindIII T. aquaticus DNA insert subcloned into plasmid BSM13+.

Figure 2 is a restriction site map of plasmid pFC85 that contains the ~2.8 kb HindIII to Asp718 T. aquaticus DNA insert subcloned into plasmid BSM13+.

As used herein, "cell", "cell line", and "cell culture" can be used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed cells" includes the primary subject cell and cultures derived therefrom without regard for the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for procaryotes, for example, include a promoter, optionally an operator sequence, a ribosome binding site, and possibly, other as yet poorly understood sequences. Eucaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

The term "expression system" refers to DNA sequences containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed with these sequences are capable of producing the encoded proteins. In order to effect transformation, the expression system may be included on a vector; however, the relevant DNA may then also be integrated into the host chromosome.

The term "gene" as used herein refers to a DNA sequence that encodes a recoverable bioactive polypeptide or precursor. The polypeptide can be encoded by a full-length gene sequence or any portion of the coding sequence so long as the enzymatic activity is retained.

"Operably linked" refers to juxtaposition such that the normal function of the components can be performed. Thus, a coding sequence "operably linked" to control sequences refers to a configuration wherein the coding sequences can be expressed under the control of the sequences.

"Non-ionic polymeric detergents" refers to surface-active agents that have no ionic charge and that are characterized, for purposes of this invention, by their ability to stabilize the enzyme herein at a pH range of from about 3.5 to about 9.5, preferably from 4 to 8.5.

The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably more than three. Its exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be derived synthetically or by cloning.

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of four different nucleotide triphosphates and thermostable enzyme in an appropriate buffer ("buffer" includes pH, ionic strength, cofactors, etc.) and at a suitable temperature. For Taq polymerase the buffer herein preferably contains 1.5-2 mM of a magnesium salt, preferably $MgCl_2$, 150-200 μM of each nucleotide, and 1 μM of each primer, along with preferably 50 mM KCl, 10 mM Tris buffer, pH 8-8.4, and 100 $\mu g/ml$ gelatin.

The primer is preferably single-stranded for maximum efficiency in amplification, but may alternatively be double-stranded. If double-stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the thermostable enzyme. The exact lengths of the primers will depend on many factors, including temperature, source of primer and use of the method. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 nucleotides, although it may contain more or fewer nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with template.

The primers herein are selected to be "substantially" complementary to the different strands of each specific sequence to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and thereby form a template for synthesis of the extension product of the other primer. However, for detection purposes, particularly using labeled sequence-specific probes, the primers

typically have exact complementarity to obtain the best results.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes each of which cut double-stranded DNA at or near a specific nucleotide sequence.

As used herein, the term "DNA polymorphism" refers to the condition in which two or more different nucleotide sequences can exist at a particular site in DNA.

As used herein, the term "nucleotide variation in sequence" refers to any single or multiple nucleotide substitutions, deletions or insertions. These nucleotide variations may be mutant or polymorphic allele variations. Therefore, the process herein can detect single nucleotide changes in nucleic acids such as occur in β -globin genetic diseases caused by single-base mutations, additions or deletions (some β -thalassemias, sickle cell anemia, hemoglobin C disease, etc.), as well as multiple-base variations such as are involved with α -thalassemia or some β -thalassemias. In addition, the process herein can detect polymorphisms, which are not necessarily associated with a disease, but are merely a condition in which two or more different nucleotide sequences (whether having substituted, deleted or inserted nucleotide base pairs) can exist at a particular site in the nucleic acid in the population, as with HLA regions of the human genome and random polymorphisms such as mitochondrial DNA. The polymorphic sequence-specific oligonucleotide probes described in detail hereinafter may be used to detect genetic markers linked to a disease such as insulin-dependent diabetes mellitus or in forensic applications. If the nucleic acid is double-stranded, the nucleotide variation in sequence becomes a base pair variation in sequence.

The term "sequence-specific oligonucleotides" refers to oligonucleotides which will hybridize to specific sequences whether or not contained on alleles, which sequences span the base pair variation being detected and are specific for the sequence variation being detected. Depending on the sequences being analyzed, one or more sequence-specific oligonucleotides may be employed for each sequence, as described further hereinbelow.

As used herein, the term "restriction fragment length polymorphism" ("RFLP") refers to the differences among individuals in the lengths of restriction fragments formed by digestion with a particular restriction endonuclease.

As used herein, the term "thermostable enzyme" refers to an enzyme which is stable to heat and is heat resistant and catalyzes (facilitates) combination of the nucleotides in the proper manner to form the primer extension products that are complementary to each nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and will proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be a thermostable enzyme, however, which initiates synthesis at the 5' end and proceeds in the other direction, using the same process as described above.

The thermostable enzyme herein must satisfy a single criterion to be effective for the amplification reaction, i.e., the enzyme must not become irreversibly denatured (inactivated) when subjected to the elevated temperatures for the time necessary to effect denaturation of double-stranded nucleic acids. Irreversible denaturation for purposes herein refers to permanent and complete loss of enzymatic activity. The heating conditions necessary for denaturation will depend, e.g., on the buffer salt concentration and the length and nucleotide composition of the nucleic acids being denatured, but typically range from about 90 to about 105°C for a time depending mainly on the temperature and the nucleic acid length, typically about 0.5 to four minutes. Higher temperatures may be tolerated as the buffer salt concentration and/or GC composition of the nucleic acid is increased. Preferably, the enzyme will not become irreversibly denatured at about 90-100°C.

The thermostable enzyme herein preferably has an optimum temperature at which it functions that is higher than about 40°C, which is the temperature below which hybridization of primer to template is promoted, although, depending on (1) magnesium and salt concentrations and (2) composition and length of primer, hybridization can occur at higher temperature (e.g., 45-70°C). The higher the temperature optimum for the enzyme, the greater the specificity and/or selectivity of the primer-directed extension process. However, enzymes that are active below 40°C, e.g., at 37°C, are also within the scope of this invention provided they are heat-stable. Preferably, the optimum temperature ranges from about 50 to 90°C, more preferably 60-80°C.

The thermostable enzyme herein may be obtained from any source and may be a native or recombinant protein. Examples of enzymes that have been reported in the literature as being resistant to heat include heat-stable polymerases, such as, e.g., polymerases extracted from the thermophilic bacteria Thermus flavus, Thermus ruber, Thermus thermophilus, Thermus aquaticus, Thermus lacteus and Thermus rubens.

The preferred thermostable enzyme herein is a DNA polymerase isolated from Thermus aquaticus. Various strains thereof are available from the American Type Culture Collection, Rockville, Maryland, and is described by T.D. Brock, J. Bact. (1969) 98:289-297, and by T. Oshima, Arch. Microbiol. (1978) 117: 189-196. One of these preferred strains is strain YT-1.

For recovering the native protein the cells are grown using any suitable technique. One such technique is described by Kaledin et al., Biokhimiya (1980), *supra*. Briefly, the cells are grown on a medium, in one liter, of nitrilotriacetic acid (100 mg), tryptone (3 g), yeast extract (3 g), succinic acid (5 g), sodium sulfite (50 mg), riboflavin (1 mg), K_2HPO_4 (522 mg), $MgSO_4$ (480 mg), $CaCl_2$ (222 mg), NaCl (20 mg), and trace elements. The pH of the medium is adjusted to 8.0 ± 0.2 with KOH. The yield is increased if cultivated with vigorous aeration up to 20 g/liter of cells at a temperature of

70°C. Cells in the late logarithmic stage (determined by absorbance at 550 nm) are collected by centrifugation, washed with a buffer and stored frozen at -20°C.

In another method for growing the cells, described in Chien et al., *J. Bacteriol.* (1976), *supra*, a defined mineral salts medium containing 0.3% glutamic acid supplemented with 0.1 mg/l biotin, 0.1 mg/l thiamine, and 0.05 mg/l nicotinic acid is employed. The salts include nitrilotriacetic acid, CaSO₄, MgSO₄, NaCl, KNO₃, NaNO₃, ZnSO₄, H₃BO₃, CuSO₄, NaMoO₄, CoCl₂, FeCl₃, MnSO₄, and Na₂HPO₄. The pH of the medium is adjusted to 8.0 with NaOH.

In the Chien et al. technique, the cells are grown initially at 75°C in a water bath shaker. On reaching a certain density, 1 liter of these cells is transferred to 16-liter carboys which are placed in hot-air incubators. Sterile air is bubbled through the cultures and the temperature maintained at 75°C. The cells are allowed to grow for 20 hours before being collected by centrifuge.

After cell growth, the isolation and purification of the enzyme take place in six stages, each of which is carried out at a temperature below room temperature, preferably about 4°C.

In the first stage or step, the cells, if frozen, are thawed, disintegrated by ultrasound, suspended in a buffer at about pH 7.5, and centrifuged.

In the second stage, the supernatant is collected and then fractionated by adding a salt such as dry ammonium sulfate. The appropriate fraction (typically 45-75% of saturation) is collected, dissolved in a 0.2 M potassium phosphate buffer preferably at pH 6.5, and dialyzed against the same buffer.

The third step removes nucleic acids and some protein. The fraction from the second stage is applied to a DEAE-cellulose column equilibrated with the same buffer as used above. Then the column is washed with the same buffer and the flow-through protein-containing fractions, determined by absorbance at 280 nm, are collected and dialyzed against a 10 mM potassium phosphate buffer, preferably with the same ingredients as the first buffer, but at a pH of 7.5.

In the fourth step, the fraction so collected is applied to a hydroxyapatite column equilibrated with the buffer used for dialysis in the third step. The column is then washed and the enzyme eluted with a linear gradient of a buffer such as 0.01 M to 0.5 M potassium phosphate buffer at pH 7.5 containing 10 mM 2-mercaptoethanol and 5% glycerine. The pooled fractions containing thermostable enzyme (e.g., DNA polymerase) activity are dialyzed against the same buffer used for dialysis in the third step.

In the fifth stage, the dialyzed fraction is applied to a DEAE-cellulose column, equilibrated with the buffer used for dialysis in the third step. The column is then washed and the enzyme eluted with a linear gradient of a buffer such as 0.01 to 0.6 M KCl in the buffer used for dialysis in the third step. Fractions with thermostable enzyme activity are then tested for contaminating deoxyribonucleases (endo- and exonucleases) using any suitable procedure. For example, the endonuclease activity may be determined electrophoretically from the change in molecular weight of phage λ DNA or supercoiled plasmid DNA after incubation with an excess of DNA polymerase. Similarly, exonuclease activity may be determined electrophoretically from the change in molecular weight of DNA after treatment with a restriction enzyme that cleaves at several sites.

The fractions determined to have no deoxyribonuclease activity are pooled and dialyzed against the same buffer used in the third step.

In the sixth step, the pooled fractions are placed on a phosphocellulose column with a set bed volume. The column is washed and the enzyme eluted with a linear gradient of a buffer such as 0.01 to 0.4 M KCl in a potassium phosphate buffer at pH 7.5. The pooled fractions having thermostable polymerase activity and no deoxyribonuclease activity are dialyzed against a buffer at pH 8.0.

The molecular weight of the dialyzed product may be determined by any technique, for example, by SDS PAGE using protein molecular weight markers. The molecular weight of one of the preferred enzymes herein, the DNA polymerase purified from *Thermus aquaticus*, is determined by the above method to be about 86,000-90,000 daltons.

The thermostable enzyme of this invention may also be produced by recombinant DNA techniques, as the gene encoding this enzyme has been cloned from *Thermus aquaticus* genomic DNA. The complete coding sequence for the *Thermus aquaticus* (Taq) polymerase can be derived from bacteriophage CH35:Taq#4-2 on an approximately 3.5 kilobase (kb) *Bgl*III-*Asp*718 (partial) restriction fragment contained within an ~18 kb genomic DNA insert fragment. This bacteriophage was deposited with the American Type Culture Collection (ATCC) on May 29, 1987 and has accession no. 40,336. Alternatively, the gene can be constructed by ligating an ~750 base pair (bp) *Bgl*III-*Hind*III restriction fragment isolated from plasmid pFC83 (ATCC 67,422 deposited May 29, 1987) to an ~2.8 kb *Hind*III-*Asp*718 restriction fragment isolated from plasmid pFC85 (ATCC 67,421 deposited May 29, 1987). The pFC83 restriction fragment comprises the amino-terminus of the Taq polymerase gene while the restriction fragment from pFC85 comprises the carboxyl-terminus. Thus, ligation of these two fragments into a correspondingly digested vector with appropriate control sequences will result in the translation of a full-length Taq polymerase.

It has also been found that the entire coding sequence of the Taq polymerase gene is not required to recover a biologically active gene product with the desired enzymatic activity. Amino-terminal deletions wherein approximately one-third of the coding sequence is absent have resulted in producing a gene product that is quite active in polymerase assays.

In addition to the N-terminal deletions, individual amino acid residues in the peptide chain comprising Taq polymerase may be modified by oxidation, reduction, or other derivatization, and the protein may be cleaved to obtain fragments that retain activity. Such alterations that do not destroy activity do not remove the DNA sequence encoding such protein from the definition of gene.

Thus, modifications to the primary structure itself by deletion, addition, or alteration of the amino acids incorporated into the sequence during translation can be made without destroying the activity of the protein. Such substitutions or other alterations result in proteins having an amino acid sequence encoded by DNA falling within the contemplated scope of the present invention.

Polyclonal antiserum from rabbits immunized with the purified 86,000-90,000 dalton polymerase of this invention was used to probe a *Thermus aquaticus* partial genomic expression library to obtain the appropriate coding sequence as described below. The cloned genomic sequence can be expressed as a fusion polypeptide, expressed directly using its own control sequences, or expressed by constructions using control sequences appropriate to the particular host used for expression of the enzyme.

Of course, the availability of DNA encoding these sequences provides the opportunity to modify the codon sequence so as to generate mutein forms also having DNA polymerase activity.

Thus, these tools can provide the complete coding sequence for Taq DNA polymerase from which expression vectors applicable to a variety of host systems can be constructed and the coding sequence expressed. It is also evident from the foregoing that portions of the Taq polymerase-encoding sequence are useful as probes to retrieve other thermostable polymerase-encoding sequences in a variety of species. Accordingly, portions of the genomic DNA encoding at least six amino acids can be replicated in *E. coli* and the denatured forms used as probes or oligodeoxyribonucleotide probes can be synthesized which encode at least six amino acids and used to retrieve additional DNAs encoding a thermostable polymerase. Because there may not be a precisely exact match between the nucleotide sequence in the *Thermus aquaticus* form and that in the corresponding portion of other species, oligomers containing approximately 18 nucleotides (encoding the six amino acid stretch) are probably necessary to obtain hybridization under conditions of sufficient stringency to eliminate false positives. The sequences encoding six amino acids would supply information sufficient for such probes.

Suitable Hosts, Control Systems and Methods

In general terms, the production of a recombinant form of Taq polymerase typically involves the following:

First, a DNA is obtained that encodes the mature (used here to include all muteins) enzyme or a fusion of the Taq polymerase to an additional sequence that does not destroy its activity or to an additional sequence cleavable under controlled conditions (such as treatment with peptidase) to give an active protein. If the sequence is uninterrupted by introns it is suitable for expression in any host. This sequence should be in an excisable and recoverable form.

The excised or recovered coding sequence is then preferably placed in operable linkage with suitable control sequences in a replicable expression vector. The vector is used to transform a suitable host and the transformed host cultured under favorable conditions to effect the production of the recombinant Taq polymerase. Optionally the Taq polymerase is isolated from the medium or from the cells; recovery and purification of the protein may not be necessary in some instances, where some impurities may be tolerated.

Each of the foregoing steps can be done in a variety of ways. For example, the desired coding sequences may be obtained from genomic fragments and used directly in appropriate hosts. The constructions for expression vectors operable in a variety of hosts are made using appropriate replicons and control sequences, as set forth below. Suitable restriction sites can, if not normally available, be added to the ends of the coding sequence so as to provide an excisable gene to insert into these vectors.

The control sequences, expression vectors, and transformation methods are dependent on the type of host cell used to express the gene. Generally, procaryotic, yeast, insect or mammalian cells are presently useful as hosts. Procaryotic hosts are in general the most efficient and convenient for the production of recombinant proteins and therefore preferred for the expression of Taq polymerase.

In the particular case of Taq polymerase, evidence indicates that considerable deletion at the N-terminus of the protein may occur under both recombinant and native conditions, and that the activity of the protein is still retained. It appears that the native proteins isolated may be the result of proteolytic degradation, and not translation of a truncated gene. The mutein produced from the truncated gene of plasmid pFC85 is, however, fully active in assays for DNA polymerase, as is that produced from DNA encoding the full-length sequence. Since it is clear that certain N-terminal shortened forms are active, the gene constructs used for expression of the polymerase may also include the corresponding shortened forms of the coding sequence.

Control Sequences and Corresponding Hosts

Procarvates most frequently are represented by various strains of *E. coli*. However, other microbial strains may also be used, such as bacilli, for example, *Bacillus subtilis*, various species of *Pseudomonas*, or other bacterial strains. In such procaryotic systems, plasmid vectors that contain replication sites and control sequences derived from a species compatible with the host are used. For example, *E. coli* is typically transformed using derivatives of pBR322, a plasmid derived from an *E. coli* species by Bolivar, et al., *Gene* (1977) 2:95. pBR322 contains genes for ampicillin and tetracycline resistance, and thus provides additional markers that can be either retained or destroyed in constructing the desired vector. Commonly used procaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the β -lactamase (penicillinase) and lactose (lac) promoter systems (Chang, et al., *Nature* (1977) 198:1056), the tryptophan (trp) promoter system (Goeddel, et al., *Nucleic Acids Res.* (1980) 8:4057) and the lambda-derived P_L promoter (Shimatake, et al., *Nature* (1981) 292:128) and N-gene ribosome binding site, which has been made useful as a portable control cassette, which comprises a first DNA sequence that is the P_L promoter operably linked to a second DNA sequence corresponding to N_{RBS} upstream of a third DNA sequence having at least one restriction site that permits cleavage within six bp 3' of the N_{RBS} sequence. Also useful is the phosphatase A (phoA) system described by Chang, et al. in European Patent Publication No. 196,864 published October 8, 1986, assigned to the same assignee. However, any available promoter system compatible with procaryotes can be used.

In addition to bacteria, eucaryotic microbes, such as yeast, may also be used as hosts. Laboratory strains of *Saccharomyces cerevisiae*, Baker's yeast, are most used, although a number of other strains are commonly available. While vectors employing the 2 micron origin of replication are illustrated (Broach, J. R., *Meth. Enz.* (1983) 101:307), other plasmid vectors suitable for yeast expression are known (see, for example, Stinchcomb, et al., *Nature* (1979) 282:39, Tschempe, et al., *Gene* (1980) 10:157 and Clarke, L., et al., *Meth. Enz.* (1983) 101:300). Control sequences for yeast vectors include promoters for the synthesis of glycolytic enzymes (Hess, et al., *J. Adv. Enzyme Reg.* (1968) 7:149; Holland, et al., *Biotechnology* (1978) 17:4900).

Additional promoters known in the art include the promoter for 3-phosphoglycerate kinase (Hitzeman, et al., *J. Biol. Chem.* (1980) 255:2073), and those for other glycolytic enzymes, such as glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters that have the additional advantage of transcription controlled by growth conditions are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and enzymes responsible for maltose and galactose utilization (Holland, *supra*).

It is also believed that terminator sequences are desirable at the 3' end of the coding sequences. Such terminators are found in the 3' untranslated region following the coding sequences in yeast-derived genes. Many of the vectors illustrated contain control sequences derived from the enolase gene containing plasmid peno46 (Holland, M. J., et al., *J. Biol. Chem.* (1981) 256:1385) or the LEU2 gene obtained from YEp13 (Broach, J., et al., *Gene* (1978) 8:121); however, any vector containing a yeast-compatible promoter, origin of replication, and other control sequences is suitable.

It is also, of course, possible to express genes encoding polypeptides in eucaryotic host cell cultures derived from multicellular organisms. See, for example, *Tissue Culture*, Academic Press, Cruz and Patterson, editors (1973). Useful host cell lines include murine myelomas N51, VERO and HeLa cells, and Chinese hamster ovary (CHO) cells. Expression vectors for such cells ordinarily include promoters and control sequences compatible with mammalian cells such as, for example, the commonly used early and late promoters from Simian Virus 40 (SV 40) (Fiers, et al., *Nature* (1978) 273:113), or other viral promoters such as those derived from polyoma, Adenovirus 2, bovine papilloma virus, or avian sarcoma viruses, or immunoglobulin promoters and heat shock promoters. A system for expressing DNA in mammalian systems using the BPV as a vector is disclosed in U.S. Patent 4,419,446. A modification of this system is described in U.S. Patent 4,601,978. General aspects of mammalian cell host system transformations have been described by Axel, U.S. Patent No. 4,399,216. It now appears, also, that "enhancer" regions are important in optimizing expression; these are, generally, sequences found upstream of the promoter region. Origins of replication may be obtained, if needed, from viral sources. However, integration into the chromosome is a common mechanism for DNA replication in eucaryotes.

Plant cells are also now available as hosts, and control sequences compatible with plant cells such as the nopaline synthase promoter and polyadenylation signal sequences (Depicker, A., et al., *J. Mol. Appl. Gen.* (1982) 1:561) are available.

Recently, in addition, expression systems employing insect cells utilizing the control systems provided by baculovirus vectors have been described (Miller, D. W., et al., in *Genetic Engineering* (1986) Setlow, J. K. et al., eds., Plenum Publishing, Vol. 8, pp. 277-297). These systems are also successful in producing Taq polymerase.

Transformations

Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described by Cohen, S. N., Proc. Natl. Acad. Sci. (USA) (1972) 69: 2110 is used for procaryotes or other cells that contain substantial cell wall barriers. Infection with Agrobacterium tumefaciens (Shaw, C. H., et al., Gene (1983) 23:315) is used for certain plant cells. For mammalian cells without such cell walls, the calcium phosphate precipitation method of Graham and van der Eb, Virology (1978) 52:546 is preferred. Transformations into yeast are carried out according to the method of Van Solingen, P., et al., J. Bact. (1977) 130:946 and Hsiao, C. L., et al., Proc. Natl. Acad. Sci. (USA) (1979) 76:3829.

Construction of a λ gt11 Expression Library

The strategy for isolating DNA encoding desired proteins, such as the Taq polymerase encoding DNA, using the bacteriophage vector lambda gt11, is as follows. A library can be constructed of EcoRI-flanked AluI fragments, generated by complete digestion of Thermus aquaticus DNA, inserted at the EcoRI site in the lambda gt11 phage (Young and Davis, Proc. Natl. Acad. Sci. USA (1983) 80:1194-1198). Because the unique EcoRI site in this bacteriophage is located in the carboxyl-terminus of the 3-galactosidase gene, inserted DNA (in the appropriate frame and orientation) is expressed as protein fused with β -galactosidase under the control of the lactose operon promoter/operator.

Genomic expression libraries are then screened using the antibody plaque hybridization procedure. A modification of this procedure, referred to as "epitope selection," uses antiserum against the fusion protein sequence encoded by the phage, to confirm the identification of hybridized plaques. Thus, this library of recombinant phages could be screened with antibodies that recognize the 86,000-90,000 dalton Taq polymerase in order to identify phage that carry DNA segments encoding the antigenic determinants of this protein.

Approximately 2×10^5 recombinant phage are screened using total rabbit Taq polymerase antiserum. In this primary screen, positive signals are detected and one or more of these plaques are purified from candidate plaques which failed to react with preimmune serum and reacted with immune serum and analyzed in some detail. To examine the fusion proteins produced by the recombinant phage, lysogens of the phage in the host Y1089 are produced. Upon induction of the lysogens and gel electrophoresis of the resulting proteins, each lysogen may be observed to produce a new protein, not found in the other lysogens, or duplicate sequences may result. Phage containing positive signals are picked; in this case, one positive plaque was picked for further identification and replated at lower densities to purify recombinants and the purified clones were analyzed by size class via digestion with EcoRI restriction enzyme. Probes can then be made of the isolated DNA insert sequences and labeled appropriately and these probes can be used in conventional colony or plaque hybridization assays described in Maniatis et al., Molecular Cloning: A Laboratory Manual (1982).

The labeled probe was used to probe a second genomic library constructed in a Charon 35 bacteriophage (Wilhelmine, A. M. et al., Gene (1983) 26:171-179). This library was made from Sau3A partial digestions of genomic Thermus aquaticus DNA and size fractionated fragments (15-20 kb) were cloned into the BamHI site of the Charon 35 phage. The probe was used to isolate phage containing DNA encoding the Taq polymerase. One of the resulting phage, designated CH35:Taq#4-2, was found to contain the entire gene sequence. Partial sequences encoding portions of the protein were also isolated.

Vector Construction

Construction of suitable vectors containing the desired coding and control sequences employs standard ligation and restriction techniques that are well understood in the art. Isolated plasmids, DNA sequences, or synthesized oligonucleotides are cleaved, tailored, and religated in the form desired.

Site-specific DNA cleavage is performed by treating with the suitable restriction enzyme (or enzymes) under conditions that are generally understood in the art, and the particulars of which are specified by the manufacturer of these commercially available restriction enzymes. See, e.g., New England Biolabs, Product Catalog. In general, about 1 μ g of plasmid or DNA sequence is cleaved by one unit of enzyme in about 20 μ l of buffer solution; in the examples herein, typically an excess of restriction enzyme is used to ensure complete digestion of the DNA substrate. Incubation times of about one hour to two hours at about 37°C are workable, although variations can be tolerated. After each incubation, protein is removed by extraction with phenol/chloroform, and may be followed by ether extraction, and the nucleic acid recovered from aqueous fractions by precipitation with ethanol. If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general description of size separations is found in Methods in Enzymology (1980) 65:499-560.

Restriction-cleaved fragments may be blunt-ended by treating with the large fragment of E. coli DNA polymerase I (Klenow) in the presence of the four deoxynucleotide triphosphates (dNTPs) using incubation times of about 15 to

25 minutes at 20 to 25°C in 50 mM Tris pH 7.6, 50 mM NaCl, 10 mM MgCl₂, 10 mM DTT and 50-100 µM dNTPs. The Klenow fragment fills in at 5' sticky ends, but chews back protruding 3' single strands, even though the four dNTPs are present. If desired, selective repair can be performed by supplying only one of the, or selected, dNTPs within the limitations dictated by the nature of the sticky ends. After treatment with Klenow, the mixture is extracted with phenol/
 5 chloroform and ethanol precipitated. Treatment under appropriate conditions with S1 nuclease results in hydrolysis of any single-stranded portion.

Synthetic oligonucleotides may be prepared using the triester method of Matteucci, et al., (*J. Am. Chem. Soc.* (1981) 103:3185-3191) or using automated synthesis methods. Kinasing of single strands prior to annealing or for labeling is achieved using an excess, e.g., approximately 10 units of polynucleotide kinase to 1 nM substrate in the
 10 presence of 50 mM Tris, pH 7.6, 10 mM MgCl₂, 5 mM dithiothreitol, 1-2 mM ATP. If kinasing is for labeling of probe, the ATP will contain high specific activity γ-³²P.

Ligations are performed in 15-30 µl volumes under the following standard conditions and temperatures: 20 mM Tris-Cl pH 7.5, 10 mM MgCl₂, 10 mM DTT, 33 µg/ml BSA, 10 mM-50 mM NaCl, and either 40 µM ATP, 0.01-0.02 (Weiss) units T4 DNA ligase at 0°C (for "sticky end" ligation) or 1 mM ATP, 0.3-0.6 (Weiss) units T4 DNA ligase at 14°C (for
 15 "blunt end" ligation). Intermolecular "sticky end" ligations are usually performed at 33-100 µg/ml total DNA concentrations (5-100 nM total end concentration). Intermolecular blunt end ligations (usually employing a 10-30 fold molar excess of linkers) are performed at 1 µM total ends concentration.

In vector construction employing "vector fragments", the vector fragment is commonly treated with bacterial alkaline phosphatase (BAP) in order to remove the 5' phosphate and prevent religation of the vector. BAP digestions are con-
 20 ducted at pH 8 in approximately 150 mM Tris, in the presence of Na⁺ and Mg²⁺ using about 1 unit of BAP per mg of vector at 60°C for about one hour. In order to recover the nucleic acid fragments, the preparation is extracted with phenol/chloroform and ethanol precipitated. Alternatively, religation can be prevented in vectors that have been double digested by additional restriction enzyme digestion of the unwanted fragments.

25 Modification of DNA Sequences

For portions of vectors derived from cDNA or genomic DNA that require sequence modifications, site-specific primer-directed mutagenesis is used. This technique is now standard in the art, and is conducted using a primer syn-
 30 thetic oligonucleotide complementary to a single-stranded phage DNA to be mutagenized except for limited mismatching, representing the desired mutation. Briefly, the synthetic oligonucleotide is used as a primer to direct synthesis of a strand complementary to the phage, and the resulting double-stranded DNA is transformed into a phage-supporting host bacterium. Cultures of the transformed bacteria are plated in top agar, permitting plaque formation from single cells that harbor the phage.

Theoretically, 50% of the new plaques will contain the phage having, as a single strand, the mutated form; 50%
 35 will have the original sequence. The plaques are transferred to nitrocellulose filters and the "lifts" hybridized with kinased synthetic primer at a temperature that permits hybridization of an exact match, but at which the mismatches with the original strand are sufficient to prevent hybridization. Plaques that hybridize with the probe are then picked and cultured, and the DNA is recovered.

40 Verification of Construction

In the constructions set forth below, correct ligations for plasmid construction are confirmed by first transforming *E. coli* strain MM294, or other suitable host, with the ligation mixture. Successful transformants are selected by amp-
 45 icillin, tetracycline or other antibiotic resistance or using other markers, depending on the mode of plasmid construction, as is understood in the art. Plasmids from the transformants are then prepared according to the method of Clewell, D. B., et al., *Proc. Natl. Acad. Sci. (USA)* (1969) 62:1159, optionally following chloramphenicol amplification (Clewell, D. B., *J. Bacteriol.* (1972) 110:667). The isolated DNA is analyzed by restriction and/or sequenced by the dideoxy method of Sanger, F., et al., *Proc. Natl. Acad. Sci. (USA)* (1977) 74:5463 as further described by Messing, et al., *Nucleic Acids Res.* (1981) 9:309, or by the method of Maxam, et al., *Methods in Enzymology* (1980) 65:499.

50 Host Exemplified

Host strains used in cloning and expression herein are as follows:

For cloning and sequencing, and for expression of constructions under control of most bacterial promoters, *E. coli*
 55 strain MM294 obtained from *E. coli* Genetic Stock Center GCSC #6135, was used as the host. For expression under control of the P_L_{RBS} promoter, *E. coli* strain K12 MC1000 lambda lysogen, N₇N₅₃cl857 SusP₈₀, ATCC 39531 may be used. Used herein is *E. coli* DG116, which was deposited with ATCC (ATCC 53606) on April 7, 1987.

For M13 phage recombinants, *E. coli* strains susceptible to phage infection, such as *E. coli* K12 strain DG98, are

employed. The DG98 strain has been deposited with ATCC July 13, 1984 and has accession number 39768.

Mammalian expression can be accomplished in COS-7 COS-A2, CV-1, and murine cells, and insect cell based expression in Spodoptera frugiperda).

5 Stabilization of Enzyme Activity

For long-term stability, the enzyme herein must be stored in a buffer that contains one or more non-ionic polymeric detergents. Such detergents are generally those that have a molecular weight in the range of approximately 100 to 250,000, preferably about 4,000 to 200,000 daltons and stabilize the enzyme at a pH of from about 3.5 to about 9.5 preferably from about 4 to 8.5 Examples of such detergents include those specified on pages 295-298 of McCutcheon's
10 Emulsifiers & Detergents. North American edition (1983), published by the McCutcheon Division of MC Publishing Co., 175 Rock Road, Glen Rock, NJ (USA), the entire disclosure of which is incorporated herein by reference. Preferably, the detergents are selected from the group comprising ethoxylated fatty alcohol ethers and lauryl ethers, ethoxylated alkyl phenols, octylphenoxy polyethoxy ethanol compounds, modified oxyethylated and/or oxypropylated straight-chain
15 alcohols, polyethylene glycol monooleate compounds, polysorbate compounds, and phenolic fatty alcohol ethers. More particularly preferred are Tween 20, from ICI Americas Inc., Wilmington, DE, which is a polyoxyethylated (20) sorbitan monolaurate, and Iconol™ NP-40, from BASF Wyandotte Corp. Parsippany, NJ, which is an ethoxylated alkyl phenol (nonyl).

The thermostable enzyme of this invention may be used for any purpose in which such enzyme is necessary or
20 desirable. In a particularly preferred embodiment, the enzyme herein is employed in the amplification protocol set forth below.

Amplification Protocol

25 The amplification protocol using the enzyme of the present invention may be the process for amplifying existing nucleic acid sequences that is disclosed and claimed in European Pat. Pub. No. 200,362, supra. Preferably, however, the enzyme herein is used in the amplification process described below.

In general, the amplification process involves a chain reaction for producing, in exponential quantities relative to the number of reaction steps involved, at least one specific nucleic acid sequence given (a) that the ends of the required
30 sequence are known in sufficient detail that oligonucleotides can be synthesized which will hybridize to them, and (b) that a small amount of the sequence is available to initiate the chain reaction. The product of the chain reaction will be a discrete nucleic acid duplex with termini corresponding to the ends of the specific primers employed.

Any nucleic acid sequence, in purified or nonpurified form, can be utilized as the starting nucleic acid(s), provided it contains or is suspected to contain the specific nucleic acid sequence desired. Thus, the process may employ, for
35 example, DNA or RNA, including messenger RNA, which DNA or RNA may be single-stranded or double-stranded. In addition, a DNA-RNA hybrid which contains one strand of each may be utilized. A mixture of any of these nucleic acids may also be employed, or the nucleic acids produced from a previous amplification reaction herein using the same or different primers may be so utilized. The specific nucleic acid sequence to be amplified may be only a fraction of a larger molecule or can be present initially as a discrete molecule, so that the specific sequence constitutes the entire
40 nucleic acid.

It is not necessary that the sequence to be amplified be present initially in a pure form; it may be a minor fraction of a complex mixture, such as a portion of the β -globin gene contained in whole human DNA (as exemplified in Saiki et al., Science, 230, 1530-1534 (1985)) or a portion of a nucleic acid sequence due to a particular microorganism which
45 organism might constitute only a very minor fraction of a particular biological sample. The starting nucleic acid sequence may contain more than one desired specific nucleic acid sequence which may be the same or different. Therefore, the amplification process is useful not only for producing large amounts of one specific nucleic acid sequence, but also for amplifying simultaneously more than one different specific nucleic acid sequence located on the same or different nucleic acid molecules.

The nucleic acid(s) may be obtained from any source, for example, from plasmids such as pBR322, from cloned
50 DNA or RNA, or from natural DNA or RNA from any source, including bacteria, yeast, viruses, organelles, and higher organisms such as plants or animals. DNA or RNA may be extracted from blood, tissue material such as chorionic villi, or amniotic cells by a variety of techniques such as that described by Maniatis et al., supra, p. 280-281.

If probes are used which are specific to a sequence being amplified and thereafter detected, the cells may be directly used without extraction of the nucleic acid if they are suspended in hypotonic buffer and heated to about
55 90-100°C, until cell lysis and dispersion of intracellular components occur, generally 1 to 15 minutes. After the heating step the amplification reagents may be added directly to the lysed cells.

Any specific nucleic acid sequence can be produced by the amplification process. It is only necessary that a sufficient number of bases at both ends of the sequence be known in sufficient detail so that two oligonucleotide primers

can be prepared which will hybridize to different strands of the desired sequence and at relative positions along the sequence such that an extension product synthesized from one primer, when it is separated from its template (complement), can serve as a template for extension of the other primer into a nucleic acid sequence of defined length. The greater the knowledge about the bases at both ends of the sequence, the greater can be the specificity of the primers for the target nucleic acid sequence, and thus the greater the efficiency of the process.

It will be understood that the word "primer" as used hereinafter may refer to more than one primer, particularly in the case where there is some ambiguity in the information regarding the terminal sequence(s) of the fragment to be amplified. For instance, in the case where a nucleic acid sequence is inferred from protein sequence information, a collection of primers containing sequences representing all possible codon variations based on degeneracy of the genetic code will be used for each strand. One primer from this collection will be homologous with the end of the desired sequence to be amplified.

The oligonucleotide primers may be prepared using any suitable method, such as, for example, the phosphotriester and phosphodiester methods described above, or automated embodiments thereof. In one such automated embodiment, diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage et al., *Tetrahedron Letters* (1981), 22:1859-1862. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Patent No. 4,458,066. It is also possible to use a primer which has been isolated from a biological source (such as a restriction endonuclease digest).

The specific nucleic acid sequence is produced by using the nucleic acid containing that sequence as a template. The first step involves contacting each nucleic acid strand with four different nucleotide triphosphates and one oligonucleotide primer for each different nucleic acid sequence being amplified or detected. If the nucleic acids to be amplified or detected are DNA, then the nucleotide triphosphates are dATP, dCTP, dGTP and TTP.

The nucleic acid strands are used as a template for the synthesis of additional nucleic acid strands. This synthesis can be performed using any suitable method. Generally it occurs in a buffered aqueous solution, preferably at a pH of 7-9, most preferably about 8. Preferably, a molar excess (for cloned nucleic acid, usually about 1000:1 primer:template, and for genomic nucleic acid, usually about 10^6 :1 primer:template) of the two oligonucleotide primers is added to the buffer containing the separated template strands. It is understood, however, that the amount of complementary strand may not be known if the process herein is used for diagnostic applications, so that the amount of primer relative to the amount of complementary strand cannot be determined with certainty. As a practical matter, however, the amount of primer added will generally be in molar excess over the amount of complementary strand (template) when the sequence to be amplified is contained in a mixture of complicated long-chain nucleic acid strands. A large molar excess is preferred to improve the efficiency of the process.

Preferably the concentration of nucleotide triphosphates is 150-200 μ M each in the buffer for amplification and $MgCl_2$ is present in the buffer in an amount of 1.5-2 mM to increase the efficiency and specificity of the reaction.

The resulting solution is then treated according to whether the nucleic acids being amplified or detected are double or single-stranded. If the nucleic acids are single-stranded, then no denaturation step need be employed, and the reaction mixture is held at a temperature which promotes hybridization of the primer to its complementary target (template) sequence. Such temperature is generally from about 35°C to 65°C or more, preferably about 37-60°C for an effective time, generally one-half to five minutes, preferably one-three minutes. Preferably, 45-58°C is used for Taq polymerase and >15-mer primers to increase the specificity of primer hybridization. Shorter primers need lower temperatures.

The complement to the original single-stranded nucleic acid may be synthesized by adding one or two oligonucleotide primers thereto. If an appropriate single primer is added, a primer extension product is synthesized in the presence of the primer, the thermostable enzyme and the nucleotide triphosphates. The product will be partially complementary to the single-stranded nucleic acid and will hybridize with the nucleic acid strand to form a duplex of strands of unequal length which may then be separated into single strands as described above to produce two single separated complementary strands. Alternatively, two appropriate primers may be added to the single-stranded nucleic acid and the reaction carried out.

If the nucleic acid contains two strands, it is necessary to separate the strands of the nucleic acid before it can be used as the template. This strand separation can be accomplished by any suitable denaturing method including physical, chemical or enzymatic means. One preferred physical method of separating the strands of the nucleic acid involves heating the nucleic acid until it is completely (>99%) denatured. Typical heat denaturation involves temperatures ranging from about 90 to 105°C for times generally ranging from about 0.5 to 5 minutes. Preferably the effective denaturing temperature is 90-100°C for 0.5 to 3 minutes. Strand separation may also be induced by an enzyme from the class of enzymes known as helicases or the enzyme RecA, which has helicase activity and in the presence of riboATP is known to denature DNA. The reaction conditions suitable for separating the strands of nucleic acids with helicases are described by Kuhn B., Abdel-Monem, M. and Hoffmann-Berling H., *CSH-Quantitative Biology*, 43: 63-67 (1978), and techniques for using RecA are reviewed in C. Radding, *Ann. Rev. Genetics*, 16:405-37 (1982). The denaturation produces two separated complementary strands of equal or unequal length.